

*Antti M. Rousi*

## Essay on institutional repositories and open access

Nykypäivän tutkijalle on tarjolla useita digitaalisia palveluja ja vertaisverkkoja, jotka tuottavat näkyvyyttä tieteellisille julkaisuille. Mihin yliopistojen julkaisuarkistot tässä kentässä sijoittuvat? Mikä on yliopistojen julkaisuarkistojen tuottama lisäarvo tutkijalle, joka hyödyntää kaupallisia vertaisverkkoja, kuten suosittua ResearchGatea? Tämä essee pohtii yliopistojen julkaisuarkistojen rooleja tieteellisessä viestinnässä. Kestävä avoimuus on esseen keskeinen teema. Esseen tarkoituksena on lisätä yliopistojen julkaisuarkistojen rooleista käytävää keskustelua.

*Antti Rousi, Aalto University Library, antti.m.rous@aalto.fi*

**T**oday's researchers have a plethora of web services to choose from to disseminate their work. The significance of, for example, widely used ResearchGate should not be understated (see e.g. Flenley, 2016). What is the added value that university-operated institutional repositories (IRs) create for researchers already using commercial scholarly networks (CSNs)? This essay explores aspects of institutional repositories that may be seen to create this value. The essay is structured into four sections, which are argued to contribute to value creation. These aspects are copyright, funder compliant preservation, open learning and societal impact. A recurring theme of the essay is sustainable openness. The aim of this essay is to promote the discussion about the roles of institutional repositories in scholarly communication.

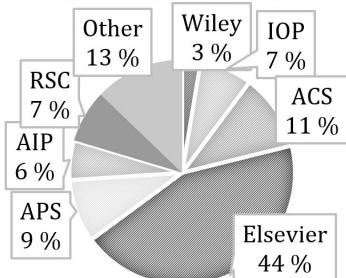
As open access (OA) refers to free and unrestricted access to scholarly content (see e.g. Laakso, 2014), such as journal articles, the term value is of interest here. The term free and unrestricted access often suggests the end-user's perspective, that is, no monetary transactions are required from the end-users for accessing OA scholarly content. The Oxford English Dictionary gives, for example, the following definitions regarding value. "The material or monetary worth of something; the amount at which something may be estimated in terms of a medium of exchange, as money or goods, or some other similar standard" and "amount of a commodity, medium of exchange, etc., which is considered to be an equivalent for something else; a fair or satisfactory equivalent or return." In their current policies the publishers of scholarly content do not require monetary transactions in regard

<b>Publisher</b>	<b>OA sharing of final version [IRs]</b>	<b>OA sharing the final version [CSNs]</b>	<b>OA sharing the accepted manuscript [IRs]</b>	<b>OA sharing the accepted manuscript [CSNs]</b>
IEEE policy source: website [1.6.2016] & email [8.12.2016]	No	No	Yes	No
IOP publishing policy source: website [2.6.2016]	No	No	Yes (After embargo)	No
Wiley policy source: website [1.6.2016]	No	No	Yes (After embargo)	No
Royal Society of Chemistry (RSC) policy source: website [1.6.2016]	No	No	Yes (After embargo)	No
American Chemical Society (ACS) policy source website [1.6.2016]	No	No	Yes (After embargo)*	No
MIT press policy source: website [1.6.2016]	Yes (After embargo)	No	Yes	No
Elsevier policy source: website [1.6.2016]	No	No	Yes (After embargo)	N/A (Subject to an agreement with the service provider)**
American Institute of Physics (AIP) policy source: website [1.6.2016]	Yes (After embargo)	No	Yes	Yes
American Physical Society (APS) policy source: website [1.6.2016]	Yes	No	Yes	Yes

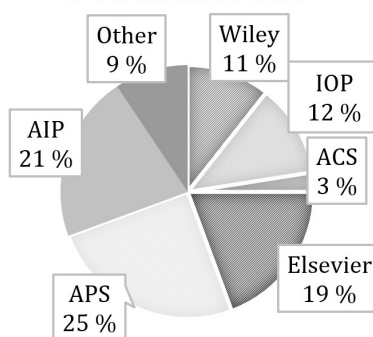
\*If mandated by the employing university \*\*No further information available at the website

Table 1 Current default publisher policies regarding sharing the post-peer review works, i.e., final publisher-prepared versions and accepted manuscripts, via IRs and CSNs. Focus in the comparison was on whether works may be shared OA on the open web and thus, e.g., private sharing with other authors or within closed groups were left out of comparison. Rights concerning sharing the publishers' paid OA options, often published under CC-license, were excluded. No retrospective comparisons concerning the policy terms were made.

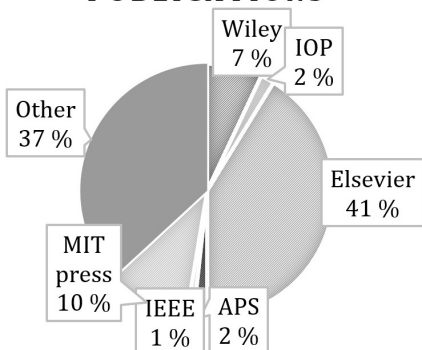
**AALTO PROF. A  
(CHEMICAL  
ENGINEERING)  
PUBLISHERS OF WEB OF  
SCIENCE LISTED  
PUBLICATIONS**



**AALTO PROF. B  
(PHYSICS)  
PUBLISHERS OF WEB OF  
SCIENCE LISTED  
PUBLICATIONS**



**AALTO PROF. C  
(NEUROSCIENCES)  
PUBLISHERS OF WEB OF  
SCIENCE LISTED  
PUBLICATIONS**



**AALTO PROF. D  
(ELECTRICAL  
ENGINEERING)  
PUBLISHERS OF WEB OF  
SCIENCE LISTED  
PUBLICATIONS**

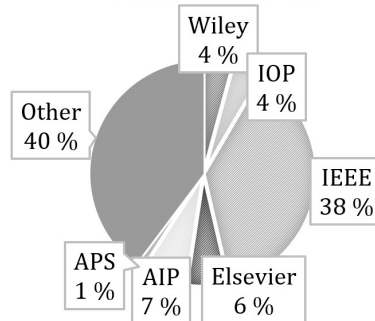


Figure 1 Publishers of Aalto University professors representing fields of chemical engineering, physics, neurosciences and electrical engineering. The examinations were based on the Thomson Reuters Web of Science publications histories of the researchers. ‘Other’ field includes publishers that the researchers less frequently worked with. Also publishing houses no longer active were included in ‘Other’.

to granting their authors the right post their works into IRs for OA (see e.g. Laakso, 2014). Thus, the transaction examined in this essay is the effort required from researchers in archiving their works into IRs and the derived return from this effort. One of the most discussed benefits or returns derived from open archiving of scholarly content is the suggested citation advantage gained from making works available OA (see Gargouri et al., 2010).

Throughout the essay comparisons between two service types are made. IRs refer to non-commercial digital document archives operated by academic institutions. IRs are designed for archiving and showcasing the research of individual academic institutions and pose no fees on accessing their content. CSNs refer to third party services that provide services to other organizations or individuals. Even if no fees on using a third-party service is posed, a service provided to, e.g., other non-commercial entities may be considered as a commercial activity (see Elsevier, 2016). In this essay, CSNs are defined as third party scholarly network services operated by private companies.

## Copyright

Previous research has suggested a citation advantage linked to OA availability of the works is linked to their accessibility in Google and Google Scholar (Gargouri et al., 2010), which implies that the advantage can be gained through any repository optimized for visibility on these search engines. There is, however, considerable difference between service types from the perspective of the copyright holders of published scholarly content. There is a substantial difference when considering author sharing rights of post peer-review works when comparing the publishers' policies on IRs and CSNs (see Table 1). In this comparison post-peer review works are defined to include both the accepted manuscript version (the terms post-print or final draft are also widely used of this version) and the final publisher-prepared version (also known as, e.g., the publisher PDF or version of record).

In the case of publishers examined in Table 1,

the current copyright policies are evidently more lenient to OA sharing of post-peer review content via IRs when compared to CSNs. All of the publishers included in the comparison allow this content to be shared via IRs. Six out of the nine publishers do not allow latter content to be shared via CSNs. One publisher states that sharing post-peer review content via CSNs could be allowed if individual service providers agree with terms set by the publisher. American Institute of Physics (AIP) and American Physical Society (APS) publishers, which allow the OA sharing of the accepted manuscripts also via CSNs, consequently allow the opening of final publisher-prepared versions via IRs. None of the publishers included in the comparison posed additional fees, such as article-processing charges, on open sharing of this content via IRs, if sharing were to comply with possible embargo terms set by the publishers. Also terminology on the policies seems to be more established regarding IRs than other web services. This often makes the work of clearing copyright issues for IR purposes less time-consuming

The aspect of copyright becomes more concrete when a university context is brought into examination. When examining the publications of Aalto University professors representing different fields (Figure 1), the potential of IRs on accumulating OA shared post-peer reviewed works becomes apparent. With only the previously compared publishers included, more than 60% of the Web of Science listed publications of Aalto professors could be shared OA as post-peer review versions from IRs with the current default policies. In two of the Aalto professor examples in Figure 1, this percentage is over 80%. As stated previously, if complying with the embargo periods set by many publishers, no additional fees would be posed on OA sharing the discussed content via IRs. Previous research has shown that instruction on copyright manners is important for the sake of establish a pro IR culture in academic institutions (Kim, 2011).

The added value by the copyright aspect regarding IRs, in contrast to CSNs, is that in taking the time to archive their works into IRs, researchers can open their post-peer review

works in compliance with the terms of the publishers (see Table 1; also Laakso, 2014). Thus, a quality of OA sharing works in compliance with the publishers' terms is derived as a return. As websites with content systematically available for the public sphere seem impossible to sustainably build on copyright infringement, this quality is seen as a prerequisite for sustainable OA sharing of scholarly content. The citation advantage gained via OA is shown to increase over time and it peaks around 6 to 7 years of openness (Gargouri et al., 2010). Thus sustainable openness is linked to gaining the full citation benefit from the OA. The above quality forms the basis for the other value creating aspects of funder compliant preservation, open learning and societal impact discussed later in this essay.

### **Funder compliant preservation**

What other aspects than complying with publishers' policies can then be seen to facilitate the sustainability of OA shared documents? First is use of proper file formats for preservation and other curating of the content. There is, for example, the PDF/A-file type that strips out the features of PDF documents whose functionalities may not be reliably preserved. Another important feature of preservation, which also enhances the citability of posted works, are persistent identifiers. Persistent identifiers can be used as permanent links in digital space. Whereas content validation and curation appears to be standard procedure incorporated into current IRs, these features, such as standardized persistent identifiers, appear not be included in current CSNs. Standardized permanent identifiers may be used to, for example, verify that a work has been shared OA via a repository, should the amount of OA published research outputs ever to be considered as a part of the university's basic funding.

Research funders are posing mandates that aim to having publicly funded research for the open circulation of knowledge to benefit both academic research and industrial uptake (see EC, 2012 and AoF, 2016). It can be argued that one of essential aspects of optimal circulation of open knowledge is sustainable openness of

research outputs. As stated previously, IRs are, both copyright and feature wise, well adapted to producing this sustainable openness. In the world of digital services 6 to 7 years, which it takes for the OA citation advantage to peak (see Gargouri et al., 2010), is a considerable time, which could exceed the lifespan of commercial services. And more, maybe even of greater importance is whether the services used for OA sharing provide the features that augment to the preservation and identification of open documents.

An aspect of IRs linked to the funder aspect is that the university has control on the metadata schema and interfaces of the repository. This relates to the ability to adjust, optimize and enrich the metadata in the landscape of changing search engines, for example. Today Google reigns, but challengers may appear. In the context of Europe, one of the most important demands on interface functionality is the compatibility with the European Commission's OpenAIRE portal. As similar interface demands may be posed by funders in the future, the control over the interfaces of the publication archive may be seen as itself an asset for the university community.

The added value of the features of IRs leading to funder compliant preservation, in contrast to CSNs, can be described as follows. Complying with the funders' demands, such as complying with Horizon2020's OpenAIRE interface, can be thought of as an important quality or return for researchers receiving funding. Also, the features as such, persistent identifiers for example, increase the citability of the works. The curation of content into file formats suitable for long-term preservation guarantees that content does not cease to be accessible due to outdated proprietary formats, for example. Another feature of IRs that may be linked to sustainable openness, is the university's control over interfaces providing open content. Funders interface requirements may vary in the future and also varied search engines may appear. A researcher taking the time to archive his/her works to IRs needs not to rely on mere commercial services in adapting to the changing landscape of digi-

tal services. In taking the time to archive their works into IRs, a quality of preservation is thus derived as a return. Along the aspect of copyright, the preservation features of IRs further augment their potential of creating sustainable openness for scholarly content.

### **Open learning**

The benefits from sustainable OA of post-peer review works may augment into more than just a citation advantage. More and more universities are launching MOOCs (Massive Open Online Courses) to open up their teaching for new audiences. The traditional license model where subscriptions to content are bought by individual universities is a poor fit with open online education. Content subscriptions may vary within universities of a country, not to mention between universities located in different continents. And more, open learning is by definition more than an affair between the universities of the world. MOOCs allow universities to showcase their teaching and research to, for example, prospective students both domestic and abroad. In the traditional model of subscription-based content most of the materials bought are unusable without a researcher or student affiliation to the university. Works shared OA are available to any eager learner with an Internet access and sufficient software to operate with most commonly used file formats.

Open online learning benefits of OA publishing in journals, particularly publishing with open licenses that allow the re-use of content in various settings (see Creative Commons, 2016). Within licensed content the option of using open licenses, Creative Commons licenses, for example, is often reserved to works, which opt to purchase openness from the publisher (hybrid OA). For the sake of OA contents efficient re-use in open learning, publishers should either specify the terms of OA shared works via IRs in relation to open online learning (to MOOCs in particular) or more frequently adapt to using open licenses.

The added value that IRs create for researchers, in contrast to CSNs, regarding the aspect of open learning stems from their potential of

creating sustainable openness. The open learning movement is still under flux and accurate returns regarding this aspect are currently hard to define. However, the combination of modern learning technology and sustainable openness of scholarly content seems to create a potential of unforeseen visibility and interaction with parties outside academia. Returns derived by researchers skillfully integrating their OA scholarly content into applications of open learning are yet to be witnessed. It seems however evident, that this skillful integration of OA content includes acknowledging the matters of copyright and preservation, which currently favor IRs over CSNs.

### **Societal impact**

The majority of research evaluation is currently focused on either citation counts or publication forum classifications of journals (e.g. JUFO). However, as Holmberg et al. (2015, 2) point out, these methods only study impact on other research publications or researchers. So called altmetric indicators track online interaction events, such as tweets, comments and blog entries, for the purpose of showcasing impact that a specific piece of research has outside academia. Even though altmetrics do not provide an alternative to, e.g., citation-based assessments of scientific impact, they are being investigated for their potential to demonstrate other kinds of societal impact. (Holmberg et al., 2015.)

The IRs and their open content seem to have potential of creating interesting dynamics into online events tracked by altmetrics. Content opened in compliance with the publishers copyright may be linked to different contexts with persistent links provided by IRs, instead of referring to subscription-based research content through DOIs provided by the publishers, for example. The dynamics of potential readership are similar to that of open learning. Whereas subscription-based content is often only available to university affiliates, OA shared content may be accessed with mere up-to-date personal computers. Links to relevant research could be posted into the comment sections of national newspapers, for example.

Here again it seems relevant that it is the post-peer review versions of research publications that may be opened from IRs. IRs could thus feed the sphere of public debate with peer-reviewed open research publications, whose quality is already examined and accepted by the scientific community. IRs could thus be used to increase the visibility of research publications in the public sphere, which in return could accumulate into altmetric indicators indicating improved visibility of research outputs outside academia. Sustainable openness of post-peer review works derived through matters of copyright and preservation seems to be of great importance again.

The added value created by IRs, in contrast to CSNs, regarding the aspect of societal impact stems also from their potential of creating sustainable openness. Post-peer review scholarly content may be opened in compliance with the publishers' terms and linked to different contexts using persistent identifiers as links, for example. The preservation features of IRs work towards having the linked works openly available for the interested reader also in the future. This sustainable open availability may further augment into improved altmetric indicators of research outputs.

## Summary

This essay built its main arguments around matters of copyright and preservation, which were seen favor IRs over CSNs. Favorable copyright and features of preservation were argued to facilitate sustainable OA of post-peer review research outputs created via IRs. Not only can sustainable OA provided by IRs work towards providing an OA citation advantage accumulating over time (see Gargouri et al., 2011), it can also work towards informing ongoing societal debates, which lead to increased visibility and use of research outputs. Sustainable OA provided by IRs may be used to fuel the current movements of open online learning and societal impact measured through altmetrics, for example. As such, IRs may be seen to complement the visibility of researchers created via CSNs and contributing to creating a stable platform for applications that leverage OA content availability.

## References

- Academy of Finland (AoF). 2016. Application guidelines – open science [online]. Available: <http://www.aka.fi/en/funding/how-to-apply/application-guidelines/open-science/> [checked 3.6.2016]
- Creative Commons. 2016. Website [online]. Available: <https://creativecommons.org/> [checked 3.6.2016]
- Elsevier. 2016. Policy on hosting articles [online]. Available: <https://www.elsevier.com/about/company-information/policies/hosting> [checked 3.6.2016]
- European Commission (EC). 2012. Background paper on open access to publications and data on Horizon2020 [online]. Available: [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/background-paper-open-access-october-2012\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/background-paper-open-access-october-2012_en.pdf) [checked 3.6.2016]
- Flenley, N. 2016. Innovations in scholarly communication – Results from the survey of Emerald authors. Emerald publishing group. Available: <http://www.emeraldgrouppublishing.com/promo/research.htm> [checked 3.6.2016]
- Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T. & Harnad, S. 2010. Self-selected or mandated, open access increases citation impact for higher quality research. *PLoS ONE*, 5(10), e13636. doi:10.1371/journal.pone.0013636
- Holmberg, K., Didegah, F., Bowman, S. Bowman, TD, & Kortelainen, T. 2015. Measuring the societal impact of open science – Presentation of a research project. *Informaatiotutkimus*, 34(4). Available: <http://ojs.tsv.fi/index.php/inf/index>
- Kim, J. 2011. Motivations of faculty self-archiving in institutional repositories. *The Journal of Academic Librarianship*, 37(3), pp. 246–254. doi: 10.1016/j.acalib.2011.02.017
- Laakso, M. 2014. Green open access policies of scholarly journal publishers: a study of what, when, and where self-archiving is allowed. *Scientometrics*, 99, pp. 475–494. doi: 10.1007/s11192-013-1205-3